Homework 2

Christophe Hunt

February 8, 2017

Contents

1 Page 69: problem 12

From this vague scenario, identify a problem you would like to study. Which variables affect the behavior you have identified in the problem identification? Which variables are the most important?

A company with a fleet of trucks faces increasing maintenance costs as the age and mileage of the trucks increase.A problem that would be interesting to study is the at what point should the truck be retired and a new vehicle purchased. The costs associated with a new purchase would need to outweigh the cost of maintaining the aged truck.

The variables of importance would be the maintenance cost as its associated with the age and mileage of the truck, any additional variables such as the severity of past repairs. I would assume that a vehicle with engine failure may have future issues until the engine is replaced, at which point the cost of maintenance may not be as severe. Additionally, the cost of a new purchase would be the continuing payments, depreciation, maintenance, fuel efficiency, and the opportunity costs of reliability.

The equilibrium of this system would be meaningful to make a data driven decision to buy a new vehicle or to keep running the aged truck.

2 Page 79: problem 11

Determine whether the data set supports the stated proportionality model.

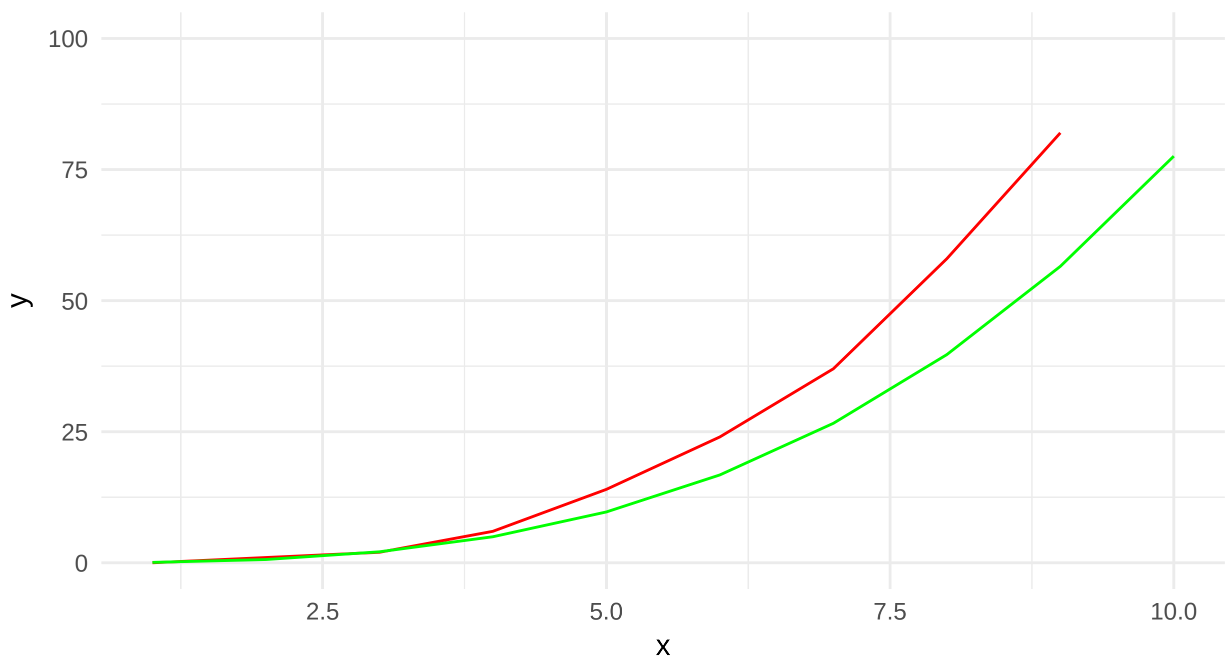
$$y \propto x^3$$

Table 1:

y	0	1	2	6	14	24	37	58	82	114
x	1	2	3	4	5	6	7	8	9	10

```
library(ggplot2)
y <- c(0, 1, 2, 6, 14, 24, 37, 58, 82, 114)
x <- c(1:10)
df <- as.data.frame(cbind(y,x))
df$y2 <- df$y^(1/3)
df$k <- df$y2/df$x
k <- mean(df$k)
df$model <- (k^3)*df$x^3

ggplot() +
  geom_line(data = df, aes(x, y), color = 'red') +
  geom_line(data = df, aes(x,model), color = 'green') +
  theme_minimal() +
  ylim(0,100)
```



The data does not support the proportion model since our data does not pass through the origin (0,0) and our slope is small comparative to 1. Our used proportional model for this data is $y = .0.078x^3$ where $k = 0.078$. This is achieved by taking $y^{\frac{1}{3}}$ and from the values in our provided data set. Then we obtain the ratio of $\frac{y^{\frac{1}{3}}}{x}$ and further obtain the mean which is 0.078. As illustrated above the model illustrates that the data does not follow the proportional model.

3 Page 94: problem 4

Lumber Cutters - Lumber cutters wish to use readily available measurements to estimate the number of board feet for lumber in a tree. Assume they measure the diameter of the tree in inches at waist height. Develop a model that predicts board feet as a function of diameter in inches.

Use the following data for your test.

Table 2:

x	17	19	20	23	25	28	32	38	39	41
y	19	25	32	57	71	113	123	252	259	294

```
x <- c(17,19,20,23,25,28,32,38,39,41)
y <- c(19,25,32,57,71,113,123,252,259,294)
df <- as.data.frame(cbind(x,y))
```

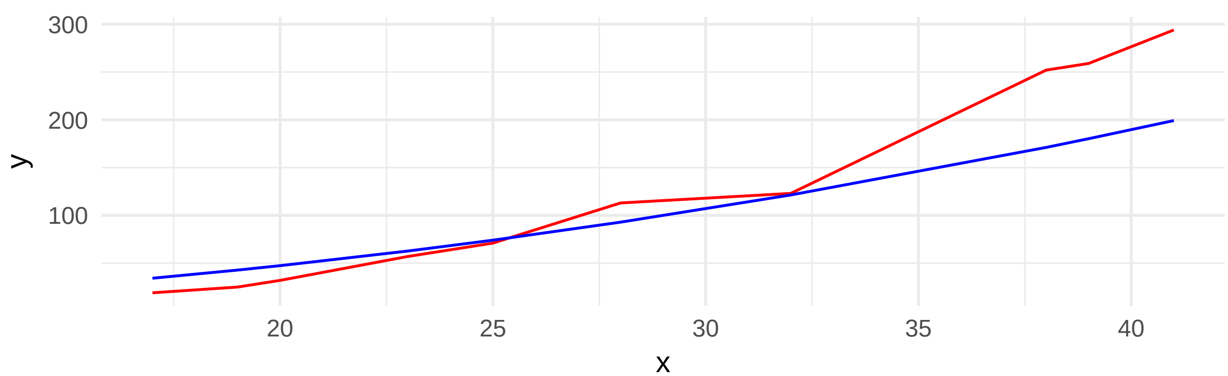
The variable x is the diameter of a ponderous pine in inches, and y is the number of board feet divided by 10.

a. Consider two separate assumptions, allowing each to lead to a model. Completely analyze each model.
b. Assume that all trees are right-circular cylinders and are approximately the same height.

We are assuming proportional change on two dimensions; the diameter change and the change in right-circular cylinders (excluding height), we therefore have the proportional model $y \propto x^2$

```
library(ggplot2)
df$y2 <- df$y^(1/2)
df$k <- df$y2/df$x
k <- mean(df$k)
df$model <- (k^2)*df$x^2

ggplot() +
  geom_line(data = df, aes(x, y), color = 'red') +
  geom_line(data = df, aes(x, model), color = 'blue') +
  theme_minimal()
```

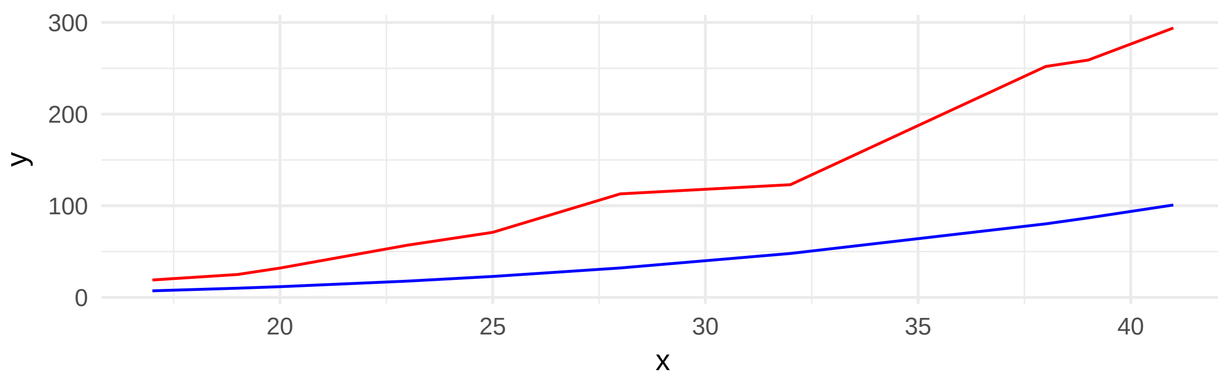


ii. Assume that all trees are right-circular cylinders and that the height of the tree is proportional to the diameter.

We are assuming proportional change on three dimensions; the diameter change, the change in right-circular cylinders, and the change in height, we therefore have the proportional model $y \propto x^3$

```
library(ggplot2)
df$y2 <- df$x^(1/3)
df$k <- df$y2/df$x
k <- mean(df$k)
df$model <- (k^3)*df$x^3

ggplot() +
  geom_line(data = df, aes(x, y), color = 'red') +
  geom_line(data = df, aes(x,model), color = 'blue') +
  theme_minimal()
```



b. Which model appears to be better? Why? Justify your conclusions.

The first model for $y \propto x^2$ because it more closely follows the shape of the observed data. It seems that the model $y \propto x^3$ does not model the growth as accuractly. This is likely due to tree growth height is not as proportional to the the diameter of the tree as one would expect. This does not seem as intuitive as I would have expected.

4 Page 99: problem 3

Discuss several factors that were completely ignored in our analysis of the gasoline mileage problem.

> Several factors that were completely ignored are air temperature and the age of the vehicle. Hybrid and electric vehicles are becoming more common and newer engines are much more fuel efficient so I believe that the age of the vehicle would have impact on an appropriate model. The ambient air temperature has impact on the air pressure of the tires and the possible friction with the road which would have an impact on our model. The model limitations statement at the end of the Automobile Gasoline Mileage is very clear that the model is quite fragile and has limit application. A limitation statement like this indicates the application of the model is very narrow.

5 Page 104: problem 2

Tests exist to measure the percentage of body fat. Assume that such tests are accurate and that a great many carefully collected data are available. You may specify any other statistic, such as waist size and height, that you would like collected. Explain how the data could be arranged to check the assumptions underlying the sub models in this section. For example, suppose the data for males between ages 17 and 21 with constant body fat and height are examined. Explain how the assumption of constant density of the inner core could be checked.

> We begin by assuming that for adults certain parts will have the same density of the inner core for different people. $D_{in} = k_1 + H + V_{in} - V_{out}$ where D_{in} = the density of the inner core and $k_1 > 0$ is the constant body fat of body parts with the same volume and density of other individuals. Also, the V_{out} is the outer volume of the core which contains the most body fat, we assume that body fat is only contained in the outer core and the inner core is muscle and bone which has different density, hence the reduction in our model. V_{in} equals the volume of the inner core.
>
> We begin by utilizing already provided submodels from the text. The volume of both the inner core and outer core is proportional to the cube of which we select to be height h. The sum of the components must be proportional to the cube of the height, or $V_{in} \propto h^3$. We can continue to use this submodel as we will assume that body fat has a proportional relationship to volume of the outer core.
>
> We will assume that body fat is represented in $V_{out} = BFtotal\ Vout$ where $BFtotal$ is the total body fat of the person and $Vout$ equals the volume of the outer core. We will further assume that this is proportional to height cubed. Substituting in the values we have the following model:
>
> $$D_{in} = k_1 + h^3 - k_2 h^3 \quad for \quad k_1, k_2 > 0$$
>
> The model suggests that variation in the denisty of the inner core is based on the volume of the inner core less the body fat k_2 and the volume of the outer core.